

Quality of care for NSAID users: development of an assessment tool

Jobanputra, Paresh

DOI:

[10.1093/rheumatology/keh566](https://doi.org/10.1093/rheumatology/keh566)

Citation for published version (Harvard):

Jobanputra, P 2005, 'Quality of care for NSAID users: development of an assessment tool', *Rheumatology*, vol. 44, no. 5, pp. 633-637. <https://doi.org/10.1093/rheumatology/keh566>

[Link to publication on Research at Birmingham portal](#)

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

Quality of care for NSAID users: development of an assessment tool

P. Jobanputra, V. Arthur, M. Pugh⁵, F. Spannuth², P. Griffiths³,
E. Thomas¹ and T. Sheeran⁴

Objective. Assessments of NSAID use based on authoritative guidelines typically overlook patients' views and nuances of medical history. Our objective was to develop an assessment tool that incorporates these aspects, and technical items, for quality of care assessments in NSAID users.

Methods. Patients newly referred to a university hospital were interviewed by a nurse using an agreed template. A multidisciplinary group of rheumatologists, nurse specialists, primary care physicians and a pharmacist reviewed current guidance and systematic reviews on NSAID use, and a series of interview transcripts. The group agreed, by informal consensus, important determinants of effective and safe NSAID use. Technical aspects of medical care and items that reflected interpersonal care were included in an index for assessing quality of care for individual patients. Interview transcripts of 100 patients were scored by panel members and reliability of scores was tested by calculating weighted percentage agreement and the κ statistic.

Results. Our final index had five domains: medical risk factors; steps taken to reduce risk; knowledge of adverse effects; NSAID dose; and cost efficiency. Each item was scored 0, 1 or 2. Scores were summed, giving a maximum of 10 (low scores indicating low quality). Intra-rater agreement was >90%; κ was 0.47–0.87 for individual domains and 0.59 for overall score. Inter-rater agreement for overall score was 95%; κ was 0.25–0.78 for domains and 0.48 for overall score. Patients with especially low scores were identified using the mode of scores for five assessors; obvious clinical concerns were identified, supporting index face validity.

Conclusions. A simple index to evaluate quality of care for NSAID users based on a patient interview is described. This may be used by one or more assessors to examine care standards and highlight deficiencies in relation to NSAID use in practice.

KEY WORDS: Quality of care, NSAIDs, Assessment, Rheumatic diseases.

Assessing the quality of medical care is especially difficult, not least because of difficulties in defining quality and care. Donabedian, in his outstanding contributions [1, 2], identified three attributes of the quality of care: structure—characteristics of health service delivery; process—activities that occur between patients and practitioners; and outcome—changes in a patient's current or future health. Francis Peabody felt that the essence of care lay 'in caring for the patient'; an aspect we refer to as interpersonal care [3].

In Britain the quality of medical care is usually assessed by auditing aspects of clinical care against a standard (or guideline). Many guidelines and quality of care measures, for example the US Arthritis Foundation's quality indicator set [4, 5], are well crafted and allow for case-mix, but have a narrow focus and tend to be concerned with technical aspects of care, such as timely delivery of drug therapy, or use of specific interventions. Guidelines and quality measures are generally concerned with the processes of care, rather than outcomes. It is generally agreed that a focus on process, for many reasons, is appropriate [1, 6]. Disadvantages of many quality measures and guidelines (other than persuasive professional scepticism [7–9]) include difficulties in accommodating important aspects of care that are particularly relevant in chronic disease: for example, eliciting patients' views

about medication [10]; allowing for variations in care because of comorbidity [11]; and other nuances of individual medical or social history, specifically the processes of caring [12, 13].

Despite wide appreciation of the benefits and potential toxicity of NSAIDs, the use of NSAIDs by many patients is suboptimal [14, 15]. In order to investigate this and to overcome limitations of guidelines on the use of NSAIDs, we set out to assess quality of care for patients with musculoskeletal pain. Our goal was to develop a tool that allowed clinicians to judge quality of care for NSAID users and that incorporated technical aspects of medical care and items that reflected interpersonal care.

Methods

Setting and patient population

Patients with musculoskeletal pain newly referred to a teaching hospital department of rheumatology were included if they agreed to be interviewed by a specialist nurse and if they were taking an NSAID (including those bought over the counter). Patients were interviewed either by telephone or face to face. In both cases interviews took place before patients saw a

Department of Rheumatology and ¹Pharmacy, Selly Oak Hospital, University Hospital Birmingham Foundation Trust, ²Kingsfield Medical Centre, ³Harborne Medical Practice, Birmingham, ⁴Cannock Chase Hospital NHS Trust, Cannock and ⁵St Mary's Hospital, Newport, Isle of Wight, UK.

Submitted 19 November 2004; revised version accepted 11 January 2005.

Correspondence to: P. Jobanputra, Department of Rheumatology, University Hospital Birmingham Foundation Trust, Selly Oak Hospital, Raddlebarn Road, Selly Oak, Birmingham B29 6JD, UK. E-mail: P.Jobanputra@bham.ac.uk

rheumatologist. Patients with any musculoskeletal pain were included, not necessarily those with a specific diagnosis. This was done to study how well NSAIDs were used in a wide variety of clinical situations since NSAIDs may be indicated for many types of skeletal or soft-tissue pain. This approach avoided the problem of diagnostic uncertainty and permitted development of an index that, potentially, had wider utility. Patients below the age of 18 yr, those with a limited knowledge of English and those who had been seen previously in our department were excluded. This study was approved by the South Birmingham Local Research Ethics Committee.

Defining data on which to base quality assessments

A multidisciplinary study group of three rheumatologists, two clinical nurse specialists, two primary care physicians and a pharmacist reviewed preliminary interview transcripts and helped develop a semistructured questionnaire for interviewing patients. The group chose items that were clinically sensible and that reflected technical aspects of care, such as previous history of peptic ulceration, and items that could reflect interpersonal care. For example, the nurse enquired whether the patient knew any adverse effects of NSAIDs and also asked if the patient was happy with their NSAID; patient responses were transcribed and quoted verbatim, if appropriate. Clinicians making quality of care judgements used this text to assess care in the relevant domain; for example, 'stomach problems', described by a patient as an example of an adverse event, may be regarded as demonstrating 'limited' knowledge by the assessor.

The final interview template (shown in Appendix 1, available at *Rheumatology* Online) sought details of medical history, such as: symptoms of dyspepsia; reasons for NSAID use; comorbidity; concomitant medications, including over-the-counter preparations; knowledge and personal experience of the benefits and hazards of NSAIDs; and functional status. The nurse also sought patients' views about how well they were coping with pain, whether NSAID therapy had been discussed with their general practitioner, what they felt about tablet-taking and whether they would be prepared to increase the dose of their NSAID if appropriate. In a small number of cases a second nurse listened to telephone interviews, with consent, in order to check the accuracy of history-taking.

Agreeing items for inclusion in quality index

The study group relied on existing guidance and systematic reviews as a starting point [16–19]; the group did not feel that an independent and exhaustive review of a well reviewed area of practice was needed. We agreed on important determinants of effective and safe NSAID use over seven 2 h meetings. It was agreed that 'goodness' of NSAID use should be based on clinical judgements and that these judgements should be made, for individual patients, in the light of an abstract of an interview obtained by the nurse interviewer. Methods for evaluating optimal NSAID use were debated extensively and a final scoring system was devised, after several incarnations, through an iterative process that included independent evaluations of a series of interview transcripts.

Testing reproducibility of our index

In order to study the reproducibility or consistency of quality judgements, 100 patients (73 women and 27 men) were interviewed. Patients were drawn from 423 consecutive patients waiting for out-patient appointments of whom 98 (23%) responded to a mailing about the study and 194 patients who were approached in clinic. Most of the excluded patients were not taking an

NSAID (180 patients) and others did not meet inclusion criteria, had already seen a specialist or declined to participate.

Data analysis

Six members of the study group—three consultant rheumatologists, two general practitioners and one pharmacist—scored each of 100 case histories independently. These 100 cases were assessed for inter-rater reliability and 30 cases, which were resubmitted randomly several weeks after the first evaluation, were rated again to study intra-rater consistency; five of the six assessors examined 30 cases twice. Case histories were submitted in batches of 10 and a minimum of 2 weeks elapsed between each batch. The original data were double-entered for accuracy. Consistency of scores is reported as the median weighted percentage of agreement and the median κ statistic, for observer pairs. Weighting of disagreements, to allow for 'partial' agreement (as three or more levels of agreement were possible in each domain of our final index) was done using the squared deviation method [20]. StatsDirect, statistical software, was used for calculations.

Results

Assessment of quality: agreed final index

The final index was designed to judge quality of care for individual NSAID users. An abstract of each patient's interview, on one A4 sheet of paper, formed the basis of quality assessments. The final index (Table 2) was modelled on Virginia Apgar's scoring method for newborn babies [21] and had five domains: medical risk factors; steps taken to reduce risk; patient knowledge of adverse effects; NSAID dose; and cost efficiency. Each domain had a scale, or score, of 0, 1 or 2. Scores for the five domains were summed, giving a maximum of 10 for each patient, higher scores reflecting better use of NSAIDs. Only brief descriptors for items in the scoring grid (Table 2) were offered and clinical judgement was demanded: more detailed descriptors proved unwieldy and increasingly complex as fresh case histories were examined and were therefore abandoned.

Two of the five items were concerned with medical risk: medical risk factors and steps taken to reduce risk. A third domain, NSAID dose, could also contribute to risk; for example, a low score would be assigned for patients taking higher than recommended doses of an NSAID. Similarly, a patient continuing to use full-dose NSAID with no perceived benefit may also attain a low score. Conversely, this domain could also contribute lower scores if an assessor judged that lower than maximum doses of NSAIDs were being used where patients had inadequate pain control and where patients, without undue risk, were prepared to use more drug.

The group felt that knowledge of adverse effects provided an indication of inter-personal care and was important for the safe use of NSAIDs. Cost efficiency was also believed to be important as it could demonstrate care in selection of drug therapy; for example, use of a proton-pump inhibitor and Arthrotec might be regarded as cost-inefficient and to constitute thoughtless drug use.

Patient characteristics

The characteristics of 100 patients whose interview transcripts were used to test the reproducibility of our index are shown in Table 1. Single NSAIDs used were: ibuprofen, 28%; diclofenac alone, 20%; celecoxib, 13%; rofecoxib, 10%; Arthrotec, 8%; and a variety of other NSAIDs, 14%. Two NSAIDs were used by 3%, and low-dose aspirin and an NSAID by 4%. Nearly half (47%)

of all patients had taken an NSAID for over a year. Over a third (35%) could not name any adverse events of NSAIDs but 53% cited potential gastrointestinal toxicity; 7% indicated that they had had no communications about the adverse effect of NSAIDs. Those who recalled information about NSAIDs cited package inserts (87%), their general practitioner (42%), the media (28%) and the internet (4%) as key sources of information.

Illustration of use of the index

We calculated the mode for each of the five domains shown in Table 2 for the 100 patients studied and whose interview transcripts had been assessed by at least five assessors. Values were then summed to give an overall quality score out of 10. This score was regarded as the gold standard assessment of quality as it relied on several expert opinions. The range of scores for these 100 patients is illustrated in Fig. 1. Four patients' (4%) cases scored less than 5 out of 10 and of these, one, a woman of 27 yr, was using two NSAIDs and had no awareness of potential adverse events. Two patients had a history of peptic ulcer or upper gastrointestinal bleeding and neither was using a proton-pump inhibitor or other appropriate gastrointestinal protective therapy, or a COX-2-selective NSAID. The fourth patient was using 225 mg of diclofenac daily and had alarming gastrointestinal symptoms, including nightly dyspepsia.

TABLE 1. Characteristics of patients studied for quality of care assessments

Characteristic (n = 100)	
Mean age (yr)	51.5 yr (range 18–75)
Sex	73 female: 27 male
Comorbidity	
None	48%
Hypertension	25%
Use of gastroprotective agents	12%
Asthma	14%
Heart disease	9%
Renal or urinary symptoms	5%
Reason for NSAID use	
Polyarticular joint pain	50%
Widespread bodily pain	33%
Pain localized to a limb or joint	10%
Spinal pain	7%
Key sources of information about adverse events	
Package inserts	87%
General practitioner	42%
Internet	4%
Other media	28%

Reproducibility of index: inter-rater and intra-rater agreement

Intra-rater agreement exceeded 90% and the κ statistic, which adjusts for agreements by chance, was 0.47–0.87 for individual domains and 0.59 for the overall score, indicating moderate or substantial agreement (Table 3). Inter-rater agreement for the overall score was 95% with a κ of 0.48, indicating moderate agreement between observers about the overall quality of care. Agreement on individual parameters of quality, however, varied and was only fair for cost efficiency ($\kappa = 0.25$) and substantial for knowledge of adverse effects ($\kappa = 0.78$).

Discussion

We describe an assessment tool to evaluate quality of care for NSAID users with musculoskeletal pain. Our index required senior clinicians to make judgements about quality of care based on a transcript of a focused patient interview. The items included in our index were agreed by informal consensus and had moderate reliability for individual assessors but, using several observers, may prove a powerful tool for quality of care evaluations in these patients. For example, by using individual domains it may be possible to create a profile for individual patients, or a population, and determine deficiencies in care that might be targeted for improvements. A quality score may also have value as a prognostic factor, for example in terms of gastrointestinal hazards, though our index was concerned with overall quality of care, not only the risks of NSAIDs.

The judgements clinicians made were primarily process of care measures, such as steps taken to reduce the risk of NSAIDs and cost efficiency. But we also included items such as patient knowledge about risks and benefits of NSAIDs, which are

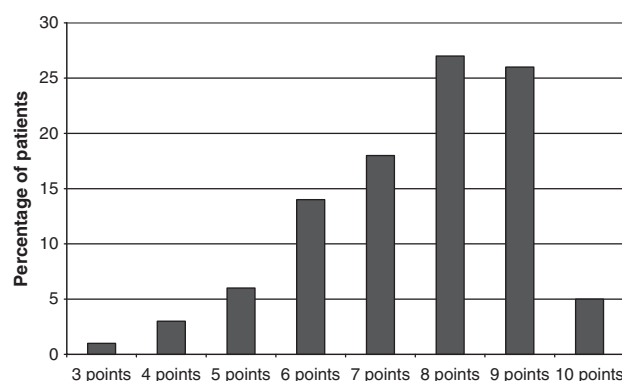


FIG. 1. Range of quality of care scores for NSAID users referred to a hospital rheumatology service (100 patients). Minimum possible score 0, maximum 10 (Table 2).

TABLE 2. Quality score for assessing NSAID use in adults*

Variable	Score		
	0	1	2
Medical risk factors	High-risk patient	Moderate risk	Low risk
Steps to reduce risk	None	Suboptimal or unknown	Optimal or not important
Knowledge of adverse effects	None apparent	Limited	Good or acceptable
NSAID dose (aspirin ≤ 150 mg/day permitted)	Overdose or 2 NSAIDs	Suboptimal for clinical situation	Optimal for clinical situation
Cost efficiency	Poor	Uncertain or intermediate	Acceptable

*In order to assign a score, the clinician makes judgements about each of the items in the first column according to the descriptors in each row and based on abstract of an individual's medical history.

TABLE 3. Reproducibility of quality of care scores: intra-rater and inter-rater agreement

	Intra-rater agreement median scores (interquartile range) 30 cases		Inter-rater agreement median scores (interquartile range) 100 cases	
	Agreement (%)	κ	Agreement (%)	κ
Medical risk factors	93 (87–95)	0.6 (0.40–0.67)	87 (84–91)	0.51 (0.43–0.62)
Steps to reduce risk	94 (94–95)	0.65 (0.60–0.77)	87 (85–91)	0.38 (0.26–0.46)
Knowledge	96 (96–98)	0.87 (0.85–0.92)	93 (93–95)	0.78 (0.75–0.80)
NSAID dose	93 (89–98)	0.47 (0.46–0.71)	90 (88–94)	0.40 (0.28–0.52)
Cost efficiency	91 (88–94)	0.63 (0.50–0.76)	82 (78–89)	0.25 (0.11–0.40)
Overall score	96 (95–96)	0.59 (0.57–0.64)	95 (95–96)	0.48 (0.44–0.52)

important outcomes. We included items that we regarded as most pertinent to quality of care for NSAID users, whilst recognizing that items included in any index for clinical assessment were, inevitably, a compromise between comprehensiveness and usability. One of the domains included, medical risk factors, reflects patient characteristics that cannot be modified by providers, though perceptions of medical risk differed (Table 3). Therefore, low scores in this domain may unfairly reduce quality of care scores. This is an important limitation of our index, but, as NSAIDs are not completely irreplaceable medicines, especially in a heterogeneous population with musculoskeletal pain, use of NSAIDs in some patients with relevant medical risk factors may be less appropriate and thus their use could reflect poorer quality of care.

The key attributes of any new index are whether the index is consistent, accurate and fit for its purpose. These attributes are often described as various forms of validity [11]. Consistency of our index was appraised by studying inter- and intra-rater reliability. The reliability of quality of care judgements was similar to evaluations of diagnostic tests such as bone scans [22] and mammograms [23]. On discussing differences, assessors readily resolved disagreements in most cases. In some cases differences were due to factual errors; for example, not reading a case history with sufficient care or not appreciating drug cost differences. The scores shown in Table 3 indicate that using several independent evaluations, or a consensus method, is likely to be more reliable for quality of care evaluations. Accuracy of our index (criterion validity) cannot be ascertained since there is no reference standard or comparable quality of care index. Similarly, formal statistical comparisons without a comparator were unfeasible and construct validity (suitability of an index for its purpose) can only be examined by further field tests and inferred from inspection of included items, which were comparable to those of the Arthritis Foundation's quality indicator set for analgesic use [5]: informing patients about risks, gastrointestinal prophylaxis, selection of NSAID, and drug monitoring.

Our index was modelled on the Apgar score [21], which remains in use after many years, and like it relies on clinical judgements which we regarded as a strength of our instrument [11]. The use of an explicit list of criteria was considered but abandoned as it became increasingly burdensome as each new patient interview, requiring yet more items, was debated. The scope of explicit quality measures, proposed by the Arthritis Foundation [5], is wider than previous guidelines on NSAID use and includes recommendations on information provision and monitoring for toxicity. Included items, however, remain focused on technical aspects of care based on if-then-because statements. For example, *if* a patient is on NSAIDs daily *and* has risk factors for developing renal insufficiency, *then* serum creatinine should be measured at baseline and at least once in the first year of treatment, *because* NSAIDs are known to cause renal impairment. Our index demanded implicit judgements from assessors and allowed the consideration of wider clinical issues; for example, quality of care could be judged where an NSAID was being taken by a patient with established renal insufficiency.

One of the strengths of our methods was the use of a multidisciplinary group and reliance on data from patient interviews rather than hypothetical situations. Health professionals were chosen from a group of colleagues with a mutual interest but did not include gastroenterologists, renal physicians or patient representatives, omissions that may be regarded as important. A disadvantage of our method was a need to interview patients to prepare a transcript for quality of care evaluations. This limits the wide applicability of our index so that large-scale evaluations of quality of care would demand significant resources: though it is possible that the index could be adapted for use with a review of medical records or a patient questionnaire. Interviews rarely took longer than 15 min, except where patients had emotional needs or communication difficulties. Inaccurate recall of medical details is an obvious limitation of interviews, but this was compensated for by gaining an insight into patient knowledge, experienced benefits and adverse effects and attitudes to the use of NSAIDs—all key parameters [24]. Interview records were not supplemented by a medical record review, though this may have improved data accuracy. However, it is clear that reliance on medical records alone underestimates quality of care [25].

Another important limitation was that we did not use formal methods of consensus to agree the items included in our index. Biases may have arisen from our informal methods [26]. For example, the differing status of health professionals, or individual personality traits, could have influenced the selection of items included or the format of our scoring method. There are, however, no ideal methods for achieving consensus and a spirit of mutual respect, many meetings, multiple opportunities for all participants to justify evaluations of clinical cases, and provision of expenses went some way to mitigating the limitations of an informal method of consensus.

In summary, our study focuses on people who use NSAIDs for musculoskeletal pain, not only those with a specific diagnosis, such as osteoarthritis. This overcomes problems of disease definition that are often ignored in guidelines and quality of care measures. We also include items that reflect elements of interpersonal care, not just those concerned with technical aspects of medical care, such as appropriate drug selection and monitoring. There are few tools to evaluate quality of care that include aspects of interpersonal care, and we believe the simple format and sufficient reliability of our index are an important step in quality of care evaluations in routine practice.

Rheumatology	Key messages
	<ul style="list-style-type: none"> A simple and novel new index for assessing quality of care for NSAID users is described.

Acknowledgements

We would like to thank, particularly, David Parry, formerly of Pfizer UK Ltd, for valuable discussions in planning and implementing this project. We would also like to thank the following for their contributions and support: Georgina Holmes and Julie Edwards, clinical nurse specialists; Sue Brailsford, research associate; the secretarial staff in the Rheumatology Unit at University Hospital Birmingham NHS Foundation Trust; and Martin Parkinson of Pfizer UK Ltd.

This study was funded by an educational grant from Pfizer UK.

Supplementary data

 Supplementary data are available at *Rheumatology* Online.

References

- Donabedian A. Explorations in quality assessment and monitoring: The definition of quality and approaches to its assessment. Vol. I. Michigan: Health Administration Press, Ann Arbor 1980.
- Donabedian A. Explorations in quality assessment and monitoring: The criteria and standards of quality. Vol. II. Michigan: Health Administration Press, Ann Arbor 1982.
- Peabody FW. Landmark article, March 19, 1927: The care of the patient. By Francis W. Peabody. *JAMA* 1984;252:813–8.
- MacLean CH, Saag KG, Solomon DH, Morton SC, Sampsel S, Klippel JH. Measuring quality in arthritis care: methods for developing the Arthritis Foundation's Quality Indicator Set. *Arthritis Rheum* 2004;51:193–202.
- Saag KG, Olivieri JJ, Patino F, Mikuls TR, Allison JJ, MacLean CH. Measuring quality in arthritis care: the Arthritis Foundation's quality indicator set for analgesics. *Arthritis Rheum* 2004;51:337–49.
- Brook RH, McGlynn EA, Cleary PD. Quality of health care. Part 2: measuring quality of care. *New Engl J Med* 1996;335:966–9.
- Cohen MD. Accentuate the positive: we are better than guidelines. *Arthritis Rheum* 1997;40:2–4.
- Shekelle PG. Why don't physicians enthusiastically support quality improvement programmes? *Qual Saf Health Care* 2002;11:6.
- Kassirer JP. The quality of care and the quality of measuring it. *New Engl J Med* 1993;329:1263–4.
- Elwyn G, Edwards A, Britten N. 'Doing prescribing': how doctors can be more effective. *BMJ* 2003;327:864–7.
- Feinstein AR. *Clinimetrics*. New Haven: Yale University Press, 1987.
- Tanenbaum SJ. What physicians know. *New Engl J Med* 1993;329:1268–71.
- Feinstein AR. Is 'quality of care' being mislabelled or mismeasured? *Am J Med* 2002;112:472–8.
- Wynne HA, Long A. Patient awareness of the adverse effects of non-steroidal anti-inflammatory drugs. *Br J Clin Pharmacol* 1996;42:253–6.
- MacGowan JR, Pandit N. The use of gastroprotection in patients using non-steroidal anti-inflammatory drugs in a general rheumatology clinic: guidelines and an audit of our practice. *Arthritis Rheum* 2001;44(Suppl.): Abstract 1591.
- American College of Rheumatology Subcommittee on Osteoarthritis Guidelines. Recommendations for the medical management of osteoarthritis of the hip and knee. *Arthritis Rheum* 2000;43:1905–15.
- Rostom A, Dube C, Wells G, Tugwell P, Welch V, Jolicœur E, McGowan J. Prevention of NSAID-induced gastroduodenal ulcers. *The Cochrane Database of Systematic Reviews* 2002, Issue 4. Article CD002296. doi:10.1002/14651858.CD002296.
- Eccles M, Freemantle N, Mason J. North of England evidence based guideline development project: summary guideline for non-steroidal anti-inflammatory drugs versus basic analgesia in treating the pain of degenerative arthritis. *BMJ* 1998;317:526–30.
- Henry D, Lim LL-Y, Garcia Roderiguez LA *et al*. Variability in risk of gastrointestinal complications with individual non-steroidal anti-inflammatory drugs: results of a collaborative meta-analysis. *BMJ* 1996;312:1563–6.
- Feinstein AR. Evaluating concordances. In: *Principles of medical statistics*. Boca Raton, FL: Chapman & Hall/CRC, 2002: 407–36.
- Apgar V. A proposal for a new method of evaluation of the newborn infant. *Curr Anaesth Anal* 1953;32:260–7.
- Whallett A, Evans N, Bradley S, Jobanputra P. Isotope bone scans: an assessment of their diagnostic use in polyarticular pain of uncertain origin. *Ann Rheum Dis* 2003;62:784–5.
- Elmore JG, Wells CK, Lee CH, Howard DH, Feinstein AR. Variability in radiologists' interpretations of mammograms. *New Engl J Med* 1994;331:1493–9.
- Fraenkel L, Wittink DR, Concato J, Fried T. Informed choice and the widespread use of anti-inflammatory drugs. *Arthritis Rheum* 2004;51:210–4.
- Luck J, Peabody JW, Dresselhaus TR, Lee M, Glassman P. How well does chart abstraction measure quality? A prospective comparison of standardized patients with the medical record? *Am J Med* 2000;108:642–9.
- Murphy MK, Black NA, Lamping DL *et al*. Consensus development methods, and their use in clinical guideline development. *Health Technol Assessment* 1998;2:1–88.